Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

A SERIES OF DIAGNOSTIC TESTS IN ARITHMETIC

WALTER S. MONROE
Director of the Bureau of Co-operative Research, Indiana University

Ideally a series of diagnostic tests on the operations of arithmetic should yield measurements of all the significant abilities within that field. It has been stated upon the basis of scientific investigations that "there are as many separate types of adding ability as there are different types of examples to be added." In the light of existing studies this general statement appears to be true not only for addition, but for all of the operations both with integers and with fractions. Thus the first step in constructing a series of diagnostic tests is the determination of the significant types of examples.

One type of example in each of the operations with integers is contained in the tables. The several combinations in an operation—for example, addition—are not equally difficult, but they call for the same type of ability: a fixed association between two written number symbols and the symbol which represents their sum. In an example such as $\begin{array}{c}7\\4\\\underline{9}\end{array}$ a different type of ability is required. The response must be given to $\begin{array}{c}7\\\underline{4,}\end{array}$ and with this partial sum 9 must be associated additively to secure the total sum. When the number of figures to be added is increased, the example is changed, but it is questionable whether a four-figure column is essentially different as to type from a five-figure column. On the other hand, there is no doubt that a twelve-figure column represents an essentially different type of example from a four- or five-figure column.

The length or span of attention is the additional factor which enters into the long-column addition.

Courtis, the author of the Standard Research Tests, has identified the following types of examples in the operations with integers:

Addition: (1) addition combinations; (2) single-column addition of three figures each; (3) "bridging the tens," as $38+7$; (4) column addition, seven figures; (5) carrying; (6) column addition with increased attention span, thirteen figures to the column; (7) addition of numbers of different lengths.

Subtraction: (1) subtraction combinations; (2) subtraction of 9 or less from a number of two digits, without "borrowing;" (3) same as the second, but with "borrowing;" (4) subtraction of numbers of two or more digits involving borrowing.

Multiplication: (1) multiplication combinations; (2) multiplicand two digits, multiplier one digit, and no carrying; (3) same as number 2, but with carrying; (4) long multiplication, without carrying; (5–8) zero difficulties; four types: $\begin{array}{r}560\\ \underline{40}\end{array}$ $\begin{array}{r}807\\ \underline{59}\end{array}$ $\begin{array}{r}617\\ \underline{508}\end{array}$ $\begin{array}{r}703\\ \underline{60}\end{array}$; (9) long multiplication with carrying.

Division: (1) division combinations; (2) simple division, no carrying; (3) same as number 2, but with carrying; (4) long division, no carrying; (5–6) zero difficulties; two cases: $\begin{array}{r}690\\ 71\overline{)48990}\end{array}$, $\begin{array}{r}302\\ 31\overline{)9362}\end{array}$; (7) long division, with carrying, "first case, the first figure of the divisor is the trial divisor and the trial quotient is the true quotient:" $\begin{array}{r}72\\ 63\overline{)4536}\end{array}$; (8) "second case, where the trial divisor is one larger than the first figure of the divisor, but the trial quotient is the true quotient:" $\begin{array}{r}63\\ 49\overline{)3087}\end{array}$; (9) "third case, where the first figure of the divisor is the trial divisor, but the true quotient is one smaller than the trial quotient:" $\begin{array}{r} 89 \\ 63\overline{)5607} \end{array}$; (10) "fourth case, where the first figure of the divisor must be increased by one to obtain a trial divisor and the second trial quotient must be increased by one to obtain a trial divisor and the second trial quotient must be increased by one to get the true quotient:" $\begin{array}{r} 79 \\ 36\overline{)2844} \end{array}$.

Except in the case of addition, where the type of example is dependent upon it, Courtis does not specify the size of the example. Thus apparently in multiplication, for example, the number of digits in both the multiplicand and the multiplier may be increased without making any essential change in the type of example. In exceptionally long examples the span of attention might become a determining factor as it does in addition. However, such examples probably have no place in the regular work of the elementary school.

In the field of common fractions Kallom[1] has analyzed the addition of two fractions and verified his analysis experimentally. The types of examples depend upon the relation existing between the denominators of the two fractions and upon the form of the answer. The denominators may be the same, in which case the fractions are called *similar*. If the denominators are different, the fractions are *dissimilar*. There are three types of dissimilar fractions: (1) "Least common denominator is the denominator of one of the fractions." (2) "Least common denominator is the product of the denominators." (3) "Least common denominator is found by factoring." The result of adding the two fractions may be in its simplest form, that is, a proper fraction and in its lowest terms. This is the *non-reducible* case. There are three *reducible* cases: (1) a proper fraction not in its lowest terms; (2) an improper fraction in its lowest terms; (3) an improper fraction not in its lowest terms. Two of the possible combinations of these conditions were found not to exist, leaving 14 types of examples in the addition of two fractions.

---

[1] *Arithmetic.* Bulletin No. 7, Department of Educational Investigation and Measurement, Boston.

For practical purposes, Kallom decided that certain types, such as 1/2+3/8 and 5/9+1/9, could be considered as included in the type 1/2+3/10. By this means he chose the following six types, which were used in the final series of tests.

ADDITION OF FRACTIONS

Test 1

(1) $\begin{array}{c}\frac{1}{4}\\ \frac{1}{4}\\ \hline\end{array}$ (2) $\begin{array}{c}\frac{3}{14}\\ \frac{1}{14}\\ \hline\end{array}$

Test 2

(1) $\begin{array}{c}\frac{1}{3}\\ \frac{1}{6}\\ \hline\end{array}$ (2) $\begin{array}{c}\frac{2}{7}\\ \frac{3}{14}\\ \hline\end{array}$

Test 3

(1) $\begin{array}{c}\frac{3}{5}\\ \frac{11}{15}\\ \hline\end{array}$ (2) $\begin{array}{c}\frac{5}{6}\\ \frac{1}{2}\\ \hline\end{array}$

Test 4

(1) $\begin{array}{c}\frac{1}{7}\\ \frac{9}{10}\\ \hline\end{array}$ (2) $\begin{array}{c}\frac{7}{9}\\ \frac{1}{4}\\ \hline\end{array}$

Test 5

(1) $\begin{array}{c}\frac{1}{10}\\ \frac{1}{6}\\ \hline\end{array}$ (2) $\begin{array}{c}\frac{4}{9}\\ \frac{5}{12}\\ \hline\end{array}$

Test 6

(1) $\begin{array}{c}\frac{1}{6}\\ \frac{9}{10}\\ \hline\end{array}$ (2) $\begin{array}{c}\frac{5}{6}\\ \frac{3}{8}\\ \hline\end{array}$

There is available no similar study of the other operations with common fractions. In subtraction all of the conditions with reference to denominators and the result exist except the case of the result in the form of an improper fraction. This would give seven types, since one combination of conditions, the reducible result with the least common denominator the product of the denominators, would not exist. Following Kallom's plan of considering certain types included in others, it appears that these four types might cover the field of subtraction:

(1) $\frac{3}{5}-\frac{3}{10}=$ (3) $\frac{3}{4}-\frac{1}{3}=$
(2) $\frac{3}{5}-\frac{4}{15}=$ (4) $\frac{5}{6}-\frac{2}{15}=$

In multiplication and division of two common fractions the relation of the denominators is not a factor, and therefore it appears that there are only four types in each operation: (1) result a proper fraction and in its lowest terms; (2) result a proper fraction but not in its lowest terms; (3) result an improper fraction but in its lowest terms; (4) result an improper fraction and not in its lowest terms. In accordance with Kallom's plan of combining types, it seems reasonable to consider that the third type may be included in the fourth.

Assuming that types can be combined as they were by Kallom in addition, we have 16 types of examples for the operations with two fractions. In addition to these, there are the types of examples involved in the addition of three or more fractions and in all of the operations with fractions, mixed numbers, and integers in their several combinations. It would seem reasonable to expect that at least 20 additional types will be found in these fields even after certain types are considered included in others.

In the field of decimal fractions there has been one study[1] to determine the types of examples which indicates "that the decimal point is not placed in quotients by means of a general rule or general ability, but by means of several specific abilities." Upon the basis of this conclusion and the types of examples in the tests used in the study it appears probable that each combination of divisors such as 5, .5 and .05 with dividends such as 16.2, 162, 1.62, and .162 constitutes a type. If this is the case, there are a number of types of examples within the range of school work, probably between 20 and 30. (It should be remembered that this statement is largely in the nature of a hypothesis based upon a few facts.)

In the field of multiplication of decimal fractions there are available no data to indicate what constitutes a type example. By analogy, we might conclude that the determination of types was similar to that in division. There is some evidence that this may not be true. In addition and subtraction there is likewise no available study of the nature of ability in the case of decimal fractions, but the most reasonable hypothesis appears to be that these two operations with decimals involve no abilities different from the abilities involved in the same operations with integers except such as occur in the writing of the numbers to be added or subtracted.

---

[1] WALTER S. MONROE, "The Ability to Place the Decimal Point in Division," *Elementary School Journal*, XVIII (December, 1917), 287-93.

It was stated in the beginning of this report that the *significant* types of examples must be determined. The analysis given above simply means that the types enumerated probably exist, but tells us nothing about their importance. That certain types, such as the addition tables, short-column addition, and the multiplication of two fractions, are socially important is an assumption that probably will not be disputed. On the other hand, some may question the social importance of some of the types of examples. For the benefit of such persons it may be pointed out that this report does not deal with the determination of the degree of the social importance of the several types enumerated. They are universally included in the course of study for arithmetic, and, for the present at least, pupils are being asked to learn to do them. It is believed that a series of tests which will effectively diagnose this field will be of service to the teacher. The determination of the types of examples which are socially important is, however, fundamental, and as reliable studies on the determination of minimum essentials become available it is not unlikely that the foregoing list of type examples will need to be modified.

The preceding survey of the operations of arithmetic indicates that there are probably not fewer than 86 significant types of examples (integers, 30; common fractions, 36; decimal fractions, 20 to 40). This total does not include the specific abilities involved in the writing and reading of numbers, in the tables of denominate numbers and in the solution of problems. A complete diagnosis of a pupil's abilities to do the operations of arithmetic would involve the measurement of his ability to do each of the 86 types of examples.

It is perhaps well to point out the meaning which the writer attaches to the measurement of the abilities to perform the operations of arithmetic. It is his thesis that measurement means the determination of the magnitude of the "dimensions" which are necessary to define the ability. If the ability is "one-dimensional," then its measurement involves the determination of only one magnitude. If it is "two-dimensional," then two magnitudes must be determined. In the case of the operations of arithmetic it is generally agreed that the abilities are "two-dimensional"—*rate* and *accuracy*. Thus the diagnosis of a pupil with respect to arithmetical abilities involves the determination of the *rate* and *accuracy* of a pupil's performance on each of the significant types of examples. In the case of a type of example, such as the combinations, where the degree of accuracy commonly attained is very high, the rate of performance becomes the more important "dimension." In his Series A, Courtis disregarded the "dimension" of accuracy for the tests upon the combinations because he found that an imperfect memorization of the fundamental number facts resulted in a reduced rate rather than in inaccurate answers. In the Cleveland Survey the number of examples done correctly was used as the measure of a pupil's ability. This is the product of the two dimensions, "rate" and "accuracy," of the pupil's work, and simplifies the handling of the scores in the survey of a large city system, but the plan probably cannot be successfully defended as the most helpful to a teacher.

Woody has attempted to measure arithmetical abilities in terms of one "dimension"—accuracy. If arithmetical abilities are essentially "*two*-dimensional," then the determination of one "dimension" is inadequate. In this connection it may be noted that Woody's procedure is in accord with conventional classroom instruction. It is customary for teachers to judge pupils with respect to the operations of arithmetic upon the basis of accuracy, disregarding almost entirely the "dimension" of rate of work.

It is obvious that "two-dimensional" measurement requires a "speed" test rather than a "power" test, because an instrument such as Woody's Arithmetic Scales yields only a "one-dimensional" measurement. Incidentally it may be remarked that it is doubtful whether this measurement is reliable.[1]

Any series of classroom tests must not require a large amount of time if it is to be used by any except the most enthusiastic workers. Thus in constructing a series of diagnostic tests in the operation of arithmetic, due regard must be had for the amount of time that will be required. Bearing this fact in mind, the writer has devised a series of 21 tests which require only 31 minutes of working time and which it is believed furnish a reasonably complete diagnosis of the operations of arithmetic, with the exception of the types of examples involving mixed numbers and integers with fractions.

The following samples from each test will illustrate the types of examples included in the several tests:

ADDITION

Test I	Test V	Test VII	Test XII	Test XV
4	7862	7	1/6+1/3 =	1/6+3/5 =
7	5013	6	5/6+1/2 =	3/12+5/8 =
2	1761	6	3/10+3/5 =	
	5872	5	5/9+2/3 =	
	3739	0		
		5		
		1		
		8		
		7		
		3		
		3		
		1		
		2		

---

[1] See WALTER S. MONROE, "An Analytical and Experimental Study of Woody's Arithmetic Scales," *School and Society*, VI (October 6, 1917), 412.

Subtraction

Test II		Test IX		Test XIII
37	94	739	1853	3/4−2/5=
5	8	367	948	5/6−3/4=

Multiplication

Test III	Test VIII	Test X			
6572	4857	560	807	617	840
6	36	37	59	508	80

Test XIV	Test XVIII		Test XX	
2/3×3/4=	657.2	67.50	487.5	57.28
2/5×3/7=	.7	.03	6.3	9.5
5/12×3/5=	46004	20250	307125	544160

In Tests XVIII and XX the pupil is simply to insert the decimal point in the product which is given. In the samples only the variations in the multiplier are given. Each multiplier is used with three types of multiplicands (657.2, 65.72, 6.572). Thus each test includes 6 types of examples.

Division

Test IV	Test VI	Test XI	Test XVI
8)3840	82)3854	47)27589	2/5÷1/3=
			4/7÷2/3=
			3/8÷2/3=

Test XIX		Test XVII		Test XXI	
.4)148	*Ans.:* 37	.03)16.2	*Ans.:* 54	.47)2758.9	*Ans.:* 587
.9)65.7	*Ans.:* 73	.07)1.82	*Ans.:* 26		
.6)1.68	*Ans.:* 28	.05).415	*Ans.:* 83	8.2)38.54	*Ans.:* 47
.7).301	*Ans.:* 43	.06)7.44	*Ans.:* 124		
				79)36.893	*Ans.:* 467

Test XI is a composite test involving the four "cases" of long division given by Courtis. Since the quotient contains three figures, there are not perfect cycles.

In Tests XVIII, XIX, and XXI the pupil is to write the answer in the proper place and insert the decimal point. In Test XXI each of the three types of divisor is placed with each of four types of dividends, thus providing 12 types of examples.

Although there are only 21 tests, 61 types of examples are represented by employing the cycle principle[1] in certain of the tests. The use of this principle lessens the definiteness of the measures, but there are two reasons which may be given for its use. In the first place, it makes possible the inclusion of a much larger number of types of examples than would otherwise be possible unless the number of tests was greatly increased. In the second place, in such tests as those on decimal fractions, variety in the types of examples is necessary for the effective measurement of the ability to place the decimal point. If all examples of a test were identical in this respect, the test would probably not be an effective measuring instrument.

The series of diagnostic tests just described were given about the middle of the year in several cities. The following report is based upon the results of giving them in three Kansas cities of the first class, the Normal Training School, and one city in Iowa. Tests 1–11, integers, were given in Grades IV–VIII; Tests 12–16, common fractions, were given in Grades V–VIII; and Tests 17–21, decimal fractions, were given in Grades VI-VIII. In Table I the median scores are based upon the scores of 300 to over 450 children in each grade. (The returns from one city were not used in computing these medians because each test was given in only one grade in that city and the scores indicated unusual conditions.) These medians may be used as tentative standards, although the number of cases is not large enough to make them very reliable standards.

A study of the facts of Table I furnished evidence of differences in the character of the arithmetical abilities required to do the examples in the different tests. Test 1 consists of examples of three digits. Two additions are required for each example, and, since 30 seconds were allowed for the test, the number of additions per minute can be obtained by multiplying the score by 4. In the same way the number of additions per minute can be computed for Test 5 (addition with carrying, five figures to the column) and Test 7 (column addition, thirteen figures to the column).

---

[1] Because of other considerations this was modified slightly in the case of the tests in common fractions.

TABLE I

MEDIAN SCORES FOR MONROE'S DIAGNOSTIC TESTS IN ARITHMETIC

The number of pupils taking the tests varies from about 300 to 450. Midyear scores. R = rate or number of examples done. A = accuracy or per cent of examples correct.

TEST NO.	GRADE									
	IV		V		VI		VII		VIII	
	R	A	R	A	R	A	R	A	R	A
1	8.3	100	8.5	100	10.2	100	12.0	100	12.7	100
2	4.4	68	5.3	100	7.1	100	8.0	100	8.9	100
3	3.6	62	4.1	67	5.1	79	5.5	83	6.2	84
4	2.2	49	2.4	55	3.4	68	4.3	79	4.6	88
5	4.2	55	4.5	59	5.3	63	5.4	63	6.1	66
6	2.4	44	2.8	57	3.5	74	4.7	70	4.5	100
7	3.6	60	4.0	70	4.5	75	5.0	78	5.4	79
8	3.2	38	3.9	60	4.9	64	5.9	68	6.5	73
9	3.9	68	5.8	100	6.6	99	7.5	100	8.5	97
10	2.8	50	3.7	65	4.8	70	5.4	73	6.6	82
11	2.0	47	2.3	48	2.7	59	3.1	64	3.4	68
12	....	....	4.5	42	6.7	63	8.5	67	9.8	73
13	....	....	4.8	49	6.3	60	7.4	65	7.8	71
14	....	....	7.6	51	10.8	70	13.0	80	13.5	75
15	....	....	5.2	40	7.1	56	8.0	61	8.5	59
16	....	....	6.2	46	9.3	81	12.0	61	8.5	59
17	....	....	....	....	3.5	47	3.9	56	4.7	54
18	....	....	....	....	8.3	100	9.5	100	11.0	100
19	....	....	....	....	4.7	52	5.3	64	6.0	59
20	....	....	....	....	8.5	100	10.0	100	11.0	100
21	....	....	....	....	3.9	44	4.3	52	4.6	52

Table II shows the number of additions per minute and the per cent of examples correct for these three tests.

Factors other than the type of example may affect the rate of adding. In Test 1 the time required to write the sum for every two additions would tend to reduce the rate. In Test 5 the time is 3 minutes, and it is possible that in the case of the younger children fatigue would decrease the rate. These two factors would tend to neutralize each other, and it is probable that the differences in rates shown in Table II cannot be explained upon any basis other than that the examples of the tests are essentially different in type and require different abilities.

There are marked differences in the per cent of examples done correctly in the three tests, but a large part of this difference is due to the difference in the size of the unit of the tests. In Test 1 the unit is two additions, and in Tests 5 and 7 the units are nineteen and twelve additions, respectively. Since an example is marked wrong if it is not entirely correct, a single incorrect addition results in two additions being counted wrong in Test 1 and nineteen in Test 5. This condition makes it impossible to compare the relative accuracy of these three tests.

TABLE II

Showing the Number of Additions per Minute and the Accuracy for Tests 1, 5, and 7

Test No.	Grade									
	IV		V		VI		VII		VIII	
	R	A	R	A	R	A	R	A	R	A
1....................	33	100	34	100	41	100	48	100	51	100
5....................	27	55	29	51	34	63	34	63	39	66
7....................	22	60	24	70	27	7	30	78	32	79

The other operations with integers are more complex, and similar comparisons cannot be made; but there is some evidence to show that the same conditions exist as in the case of addition.

The operations with common fractions are even more complex than the operations with integers. The pupil must first decide upon the procedure he is to employ before he can begin calculating. For example, he must decide how he is going to treat the numerators and denominators. Will he add the numerators for the numerator of the sum and the denominators for the denominator of the sum, as some pupils did on the test, or will he treat the denominators in a different manner? After the sum is obtained, it must be reduced to the simplest form. The pupil may neglect to do this or do it only partially. In carrying out the correct plan the pupil may, of course, make errors in calculation. Either an incorrect or incomplete plan or a mistake in calculation is counted as an error.

In Fig. 1 the distributions of the sixth-grade pupils are given for the five tests on common fractions. Unfortunately, the time allowed for these tests was such that several pupils, particularly in Test 14, did all of the examples. Disregarding this feature of the distributions and remembering that the five tests were given in succession to the same pupils, we find that the curves for the several tests present certain striking characteristics. For Test 13, subtraction, we have approximately the usual normal distribution, and in Tests 12 and 15, addition, there is a clear indication of the normal distribution. The curves for Test 14, multiplication, and Test 16, division, approximate a rectangular distribution more nearly than any other form. This form of distribution is even more conspicuous in the fifth grade, and it is evident in the seventh and eighth grades. An examination of the test papers shows that some pupils had difficulty in determining upon the plan to use, and as a result did only a few examples. Others decided upon a plan immediately (sometimes an incorrect one) and completed the test. This undoubtedly accounts for the unusual form of distribution.

In view of the fact that multiplication and division of fractions involve fewer possibilities as to types of examples, it is surprising that this condition should exist for these two operations and not for addition and subtraction. It may be that the explanation is that addition and subtraction receive more emphasis in teaching.

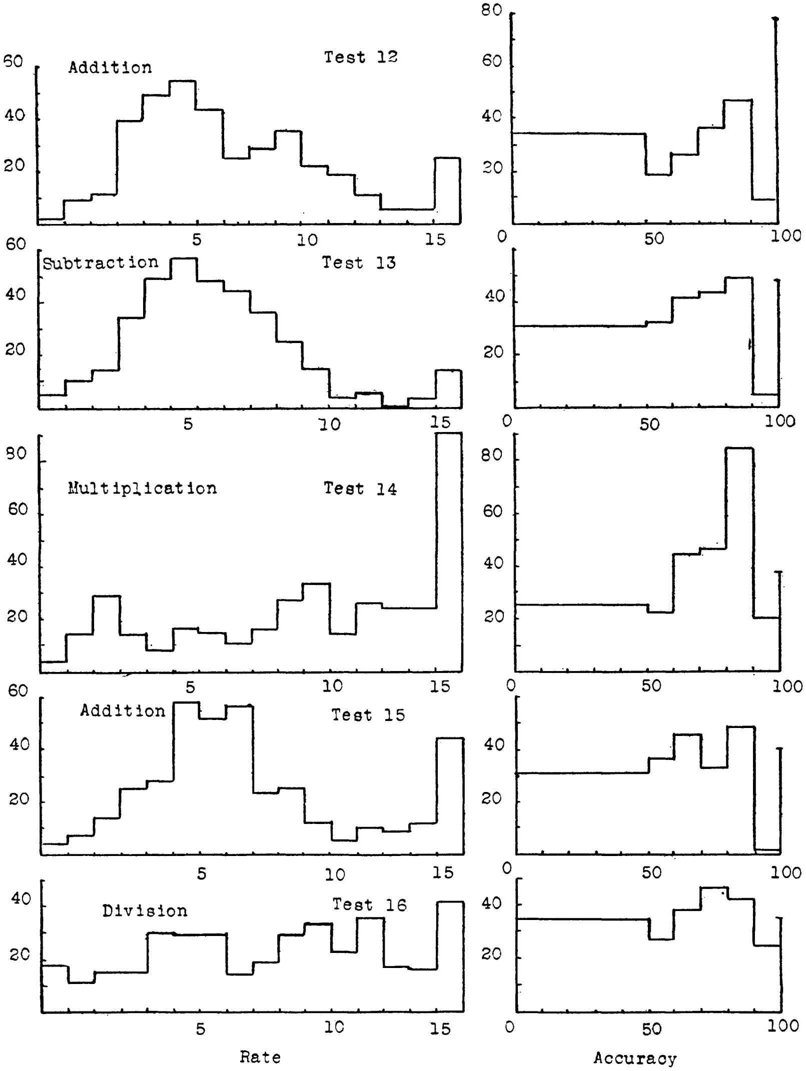

FIG. 1.—Showing distributions of sixth-grade pupils on Tests 12–16 (common fractions).

Tests 17–21 are on decimal fractions; Tests 17, 19, and 21 on division, and Tests 18 and 20 on multiplication. The most conspicuous feature of the median scores given in Table I is that in general the rate for multiplication is more than double that for division, and the accuracy is nearly double. The time allowance was the same for each test, but in the division tests it was necessary for the pupil to copy the answer. This would affect the rate, and might possibly account for the difference in rate for the two operations, but it could not account for the great difference in accuracy. Some who used the tests stated that the directions for the division tests were not clear to the pupils; but, again, it does not seem that this condition would affect the accuracy to such a large extent. The fact that there is little difference between the median scores for the several division tests indicates that failure to understand the directions would not seriously affect the scores. There seems to be no doubt that it is much easier to place the decimal point in multiplication than in division.

After studying the scores for the multiplication tests and examining some of the test papers the writer entertains some doubts concerning the soundness of his hypothesis with reference to types of examples in the multiplication of decimal fractions. It may be that there is only one essential type, but further experimental study is needed to determine this point.

The median scores of Table I slightly modified to smooth out some of the irregularities were used as a basis of the scale for Figs. 2, 3, and 4. The two scores of a test are represented graphically on the sides of an elongated rectangle, the number of examples attempted on the upper side and the number of examples right on the lower side. The scale on each line has been so chosen that the median scores of Table I for a grade fall upon a vertical line. Thus if the scores of a sixth grade were exactly standard, they would all fall on the vertical line below VI. This, of course, makes impossible a uniform scale, but the plan has the advantage of making very obvious differences between scores and standards, and, what is probably more important, it indicates the meaning of this difference. To know in terms of the differences between standards of the successive grades the absolute amount of the difference between a score and the standard is not so important as to know the meaning of the difference which exists. For example, a difference of 1.0 in Test 11 has a greater significance than a difference of 2.0 in Test 1, because the units are larger and the time allowances for the two tests do not have the same ratio as the size of the example.

In order to simplify graphical representation, the number of examples right has been used instead of the accuracy or the per cent of examples right. This number has been calculated by multiplying the number attempted by the per cent right. The direction of the slope of the line connecting the points representing the two scores of a test indicates the relative degree of accuracy. If the line is vertical, the accuracy is just standard. If the point representing the number of examples right is to the left of the point representing the rate, the accuracy is below standard. If it is to the right, the accuracy is above standard.

In Fig. 2 the median scores of one city are represented. In this city the Courtis Standard Research Tests, Series B, have been used systematically for several years, and before their appearance Series A was used. There has been a definite effort to improve the instruction in the operations of arithmetic. The solid line running from IV represents the median scores for the fourth grade. In Test 1 the fourth-grade pupils of this city are above fifth-grade standards and nearly as good as both the fifth- and sixth-grade pupils. However, in Tests 5 and 7, which are also addition tests, the fourth-grade pupils

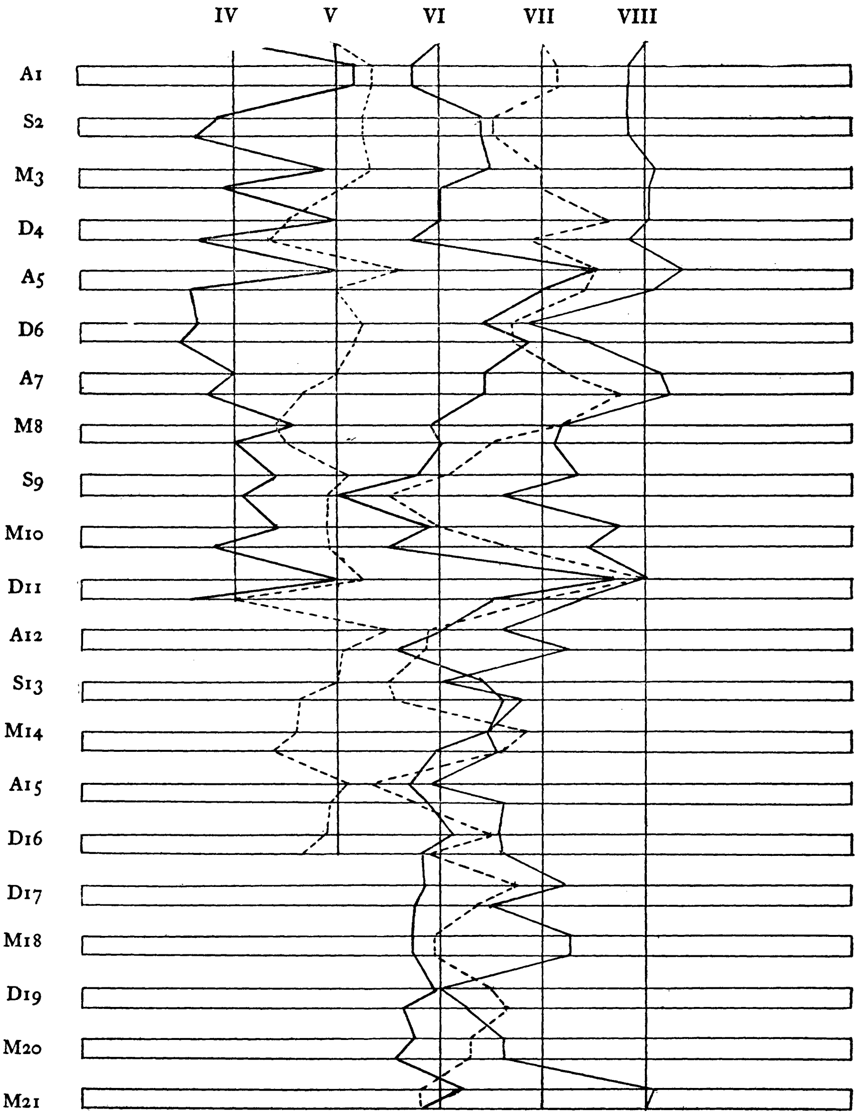

FIG. 2.—Showing the median scores for one city.

are below the fourth-grade standard for number of examples right. In the case of the sixth grade the median scores are below standard for Test 1, and above standard for Tests 5 and 7. The differences between the relative positions of the medians for the addition tests are greater than the differences between the positions of the medians for the different operations. Similar conditions exist with respect to the other operations, and this general statement seems to be justified. The differences in the abilities of the pupils of this city to do different types of examples in the same operations are as large as the differences in their abilities to do the different operations. These facts corroborate the hypothesis upon which the tests were constructed, that ability to add is not a single ability, but that there are as many different abilities as there are types of examples to be added.

It also appears from Fig. 2 that the seventh- and eighth-grade pupils are conspicuously below standard in common fractions (Tests 12–16) and in decimal fractions (Tests 17–21). One exception is the eighth-grade medians for Test 21. The existence of this condition is no doubt due to the lack of attention to the operations with fractions. This probably was caused by the attention of the teachers being focused upon the operations with integers as a result of the use of the Courtis Standard Research Tests, Series B. Thus we have a demonstration of what is likely to happen when a series of tests is used which does not cover the field of the subject. To offset the tendency to focus greater instructional effort upon those abilities which the tests measure, the limitations of the field of the series should be definitely recognized, and adequate instruction should be provided for the other phases of the subject.

In Fig. 3 there are represented the median scores for three sixth-grade classes in three different buildings in the city whose median scores are represented in Fig. 2. Classes A and

B have 33 pupils each, and Class C 23 pupils. The figure shows two significant facts: (1) certain points of non-uniformity between the medians of the three classes and (2) the non-uniform abilities of any one of the classes. In certain of the tests, such as 2, 6, 8, 15, 18, and 20, the median scores of the three classes fall within an interval of about one grade. In others, notably Tests 13 and 21, the extreme difference is much greater. Evidently the teachers of the different classes have not been placing equal emphasis upon the different types of examples. Take, for example, the addition tests (1, 5, and 7). The teacher of Class B has been placing much emphasis upon the type of example in Test 5 (short-column addition with carrying), although it should be noted that she has not neglected the other types of addition examples. On the other hand, the teacher of Class C has neglected short-column addition (Test 1), while the teacher of Class A has given much emphasis to it. In Test 7 there are represented three degrees of emphasis.

The non-uniform character of the abilities of a class is very obvious from the irregularity of the lines representing their abilities. Perfect uniformity would be represented by a straight vertical line. The non-uniformity in the abilities is due, in part at least, to the failure of the teacher to place the appropriate degree of emphasis upon the several types of examples. Some types doubtless require more emphasis than others, and it is the teacher's problem (or is it the problem of the maker of the course of study?) to determine the degree of emphasis which is needed for each type.

In Fig. 4 there are shown the scores of two sixth-grade pupils selected almost at random from Class A in Fig. 3. H. H. is a twelve-year-old boy and D. H. a girl, age not given. One should expect greater variations when dealing with the scores of individual pupils, but the variations of these scores must be surprising to one who has not studied the subject.

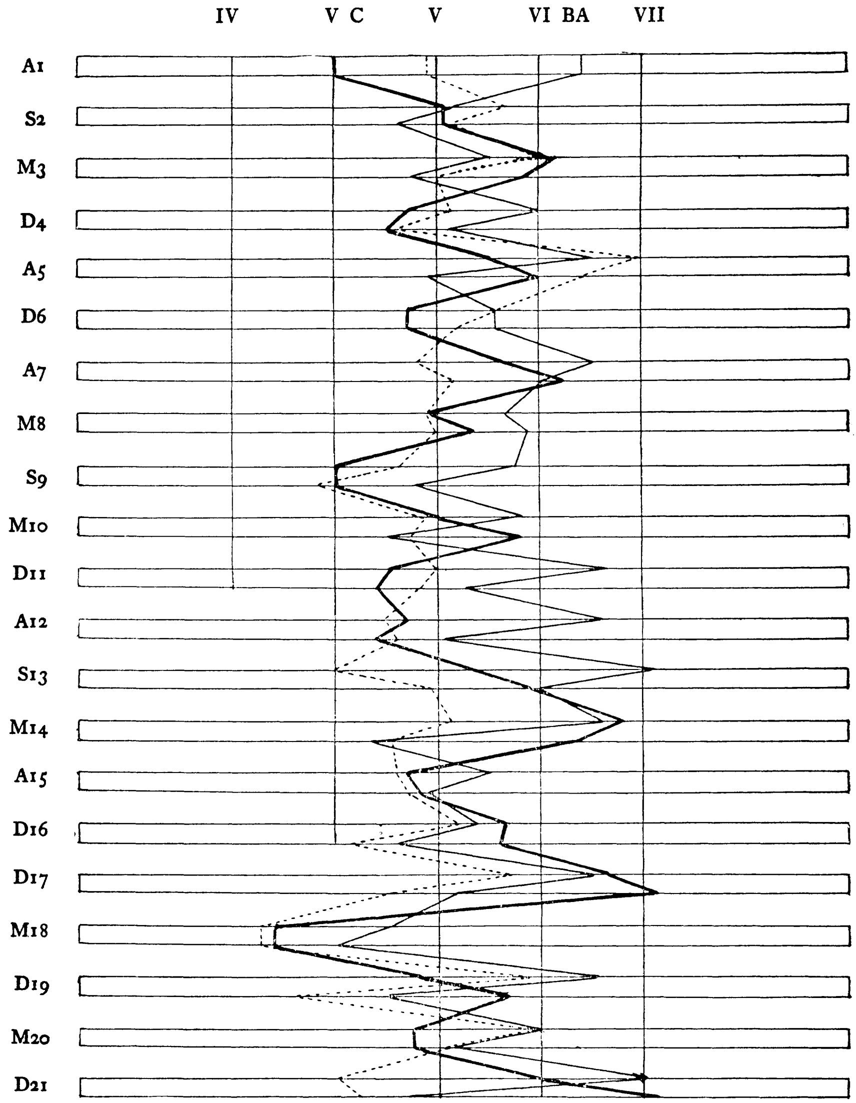

FIG. 3.—Showing the median class scores of three sixth-grade classes in the same city (Fig .2).

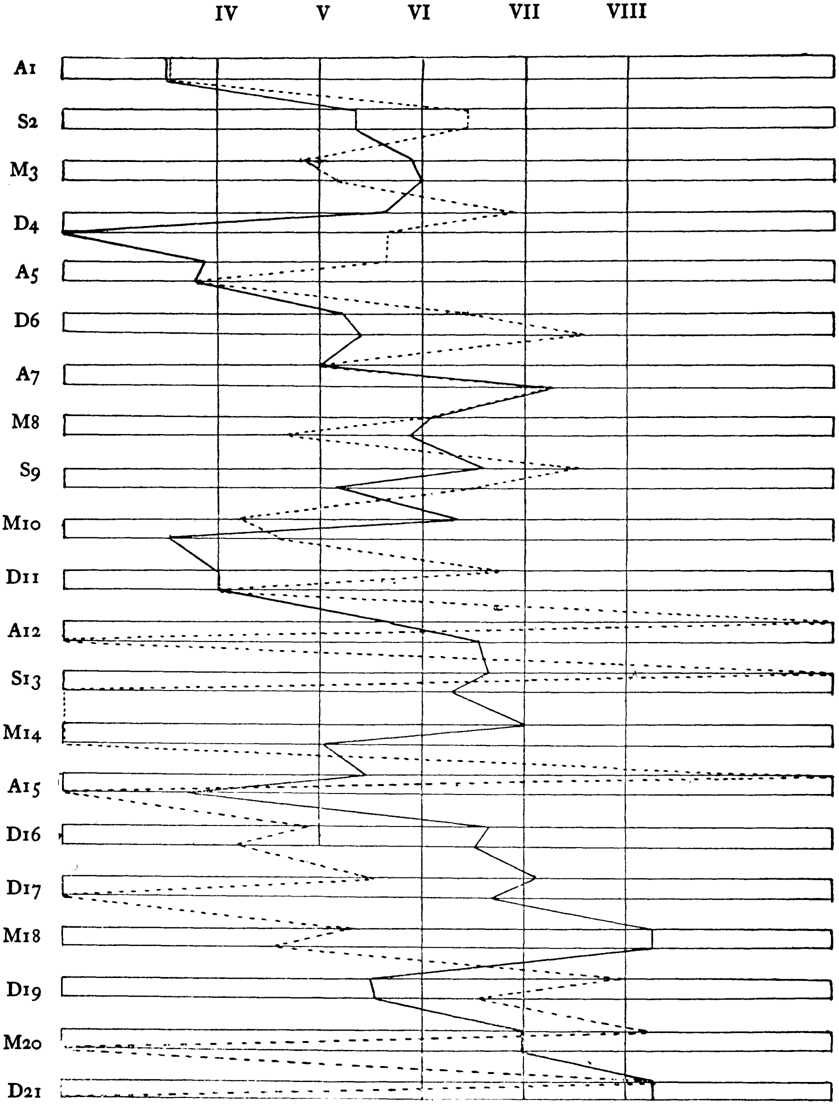

FIG. 4.—Showing the individual scores of two sixth-grade pupils in the same class.

Each of these pupils has scores on certain tests conspicuously below the standard for the fourth grade, and on other tests has scores conspicuously above eighth-grade standards.

How much of this variability is due to variation in successive performance and how much is due to actual relative differences in the several abilities, this study does not reveal. Courtis has concluded that the scores on his tests are reliable in 70 per cent of the cases. If this condition holds here, the non-uniformity of development cannot be explained by the variability of performance, and it represents a serious defect in the product of instruction in arithmetic.

It naturally occurs to one to ask if these irregularities can be eliminated. The writer has collected no data on this question for this particular series of tests. Other investigators have reported similar cases where irregular development has been dealt with satisfactorily, and there is reason to believe that much, if not all, of the irregularity shown both for classes and for individuals can be eliminated when appropriate instruction is provided.

The facts described above show that the product of arithmetical instruction is complex, much more complex than teachers and supervisors generally realize. In recent years much emphasis has been placed upon individual differences. More recently attention has been called to the fact that, although some general factors may function, ability is specific in many cases. Thus the teacher must recognize that not only do pupils of a group differ widely in ability, but also that each pupil is likely to possess each of a group of abilities in widely different degrees.

The fact that the scores obtained by using these tests show such great variations in the relative degree of ability in the different types of examples when the pupils have been measured with the Courtis Standard Research Tests, Series B, at regular intervals is evidence of the need which exists for a series of diagnostic tests. Teachers are probably failing to place the appropriate degrees of emphasis upon the different types of examples because they are ignorant of what types exist or because they do not know the degree of ability which has been attained by their class and, much less, the degree of ability attained by the individual pupils. A series of diagnostic tests such as described in this report is valuable to the teacher in two ways: first, as a statement of the important types of examples; secondly, as an instrument of diagnostic measurement.